NEWSLETTERS   PRESS   MY ACCOUNT   DONATE   CONTACTED BY US?

Read our research on: Economy & Work | Election 2024 | News Habits

Pew Research Center

Search pewresearch.org...

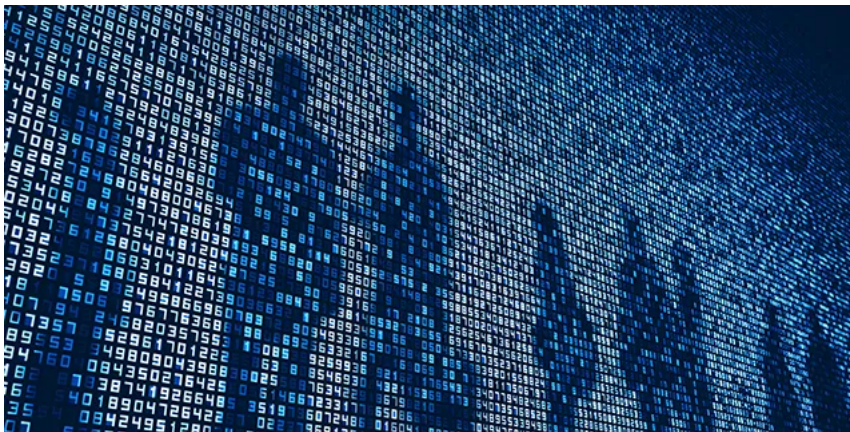RESEARCH TOPICS ▾   PUBLICATIONS   OUR METHODS   SHORT READS   TOOLS & RESOURCES   EXPERTS   ABOUT US

Home > Research Topics > Politics & Policy > U.S. Elections & Voters > Voters & Voting...

SHORT READS | FEBRUARY 15, 2018   SHARE ⬆

Q&A: The growing use of 'voter files' in studying the U.S. electorate

BY **DREW DESILVER**

(Saul Gravy/Getty Images)

Since the rise of modern survey research three-quarters of a century ago, much of what we know about voter attitudes has been based on interviews with random samples of voters, sometimes combined with tallies of actual votes and the characteristics of voters in particular locations. But relatively new technologies and public policy changes have given pollsters, academics and other researchers – not to mention candidates and political professionals – a potentially powerful new tool for probing the electorate: digital databases that claim to cover most of the U.S. adult population.

Despite their widespread use, these national databases (commonly called "voter files") are little known by the general public and haven't received much scholarly attention. Pew Research Center's Ruth Igielnik, a research associate, and Senior Survey Advisor Scott Keeter recently released an extensive study of the completeness and accuracy of commercially available voter files, which they assessed by matching participants from the Center's nationally representative American Trends Panel to five different voter files. We asked Keeter and Igielnik to talk about their work, including how voter files are built and what they're used for. Their replies have been edited for clarity and concision.

What is a "voter file" anyway? Isn't voting information private?

Good question! These digital databases – commonly known as "voter files" – are built by commercial organizations using official, publicly available government records of who is registered to vote and who cast ballots in past elections. They not only give a nationwide

RELATED

SHORT READS | AUG 28, 2024

Key things to know about U.S. election polling in 2024

FACT SHEET | JUL 8, 2024

National Public Opinion Reference Survey (NPORS)

REPORT | JUL 8, 2024

How Pew Research Center Uses Its National Public Opinion Reference Survey (NPORS)

SHORT READS | JUN 26, 2024

Q&A: What is the American Trends Panel, and how has it changed?

SHORT READS | JUN 26, 2024

How do people in the U.S. take Pew Research Center surveys, anyway?

TOPICS

Methodological Research

Political & Civic Engagement

Voter Files

Voter Participation

MOST POPULAR

1   Key things to know about U.S. election polling in 2024

2   Asian Americans: A Survey Data Snapshot

picture of voter registration and election turnout, but also usually include information from outside data sources (such as consumer data vendors, credit bureaus and political organizations) and are marketed as providing a rich and comprehensive record for nearly every American adult.

However, while information on the voter file record indicates *whether or not* someone voted in a given election, it does not indicate *whom* they voted for. That stays with you in the voting booth.

Ruth Igielnik, Pew Research Center research associate

Who uses these voter files, and how?

Beyond campaigns, voter files are used by pollsters, academic researchers and journalists who want to better understand Americans and answer specific questions about voters and elections. Some useful applications of voter files include the [demographic composition](#) of the electorate, what kinds of voters [tend to vote early or by absentee ballot](#), who [votes in primary elections](#) and what kinds of people are [habitual vs. episodic voters](#).

Scott Keeter, Pew Research Center senior survey advisor

Why did you decide that now was a good time to evaluate the quality of voter files, as opposed to, say, five or 10 years ago?

While campaigns have been using these databases for years, the recent rise in use by academics, researchers and journalists drew our attention to voter files as another potential tool in understanding the U.S. electorate. Research on voter files also fits into the Center's broader agenda to study new polling methodologies. Pollsters are increasingly using voter files as a potential sampling source for surveys, and in the future we plan to study the files' usefulness for this purpose.

You said earlier that voter files are built by commercial organizations. Do results differ much between these vendors, given that they all get their basic data from the same official sources?

It's true that official election records about registration and turnout don't vary much. What does differ is what other information gets added to the files, such as turnout in past elections and other addresses – or, in the case of our study, who gets matched from our American Trends Panel survey data. While commercial voter file companies, or vendors, all get their data from official government records, how they gather and match those records and what additional information they provide about people make up the "secret sauce" that distinguishes them from each other.

First, vendors vary in how they track Americans as they move across states. Although federal law (the [Help America Vote Act of 2002](#)) requires each state to keep an individual list of its voters, the administration of elections in the U.S. historically has been highly decentralized, so harmonizing state records can be challenging. According to June 2017 data from our American Trends Panel, 16% of Americans said they had lived at their

current address for less than one year. Those people become even harder to find if they've changed their name or moved often.

Then there's the additional analysis vendors provide. Commercial voter files typically offer scores for things like partisanship and expected turnout for future elections. These scores are generated through a process known as predictive modeling, and they vary in their availability and accuracy. For example, on the whole, models provided by the vendors in this study were better at predicting someone's party affiliation and race than their education or income.

How did you choose which voter file vendors to use for this study?

The vendors used in this study represent five of the most prominent and commonly used voter files. The vendors included ones that were traditionally nonpartisan, as well as those whose clients were mainly Democratic and politically progressive or Republican and politically conservative.

The purpose of the study, though, wasn't to focus on individual vendors or single out specific vendors as particularly good or bad. We wanted to look at a representative sample of the major commercial voter files so we could get a sense of their strengths and weaknesses with regard to supplementing existing survey data or serving as sampling frames for new public opinion surveys. Because of that, we anonymized the files and left the vendor names out of this analysis.

How much did the completeness and accuracy of the five voter files you studied vary?

A lot, depending on what you were measuring. For example, the vendors were all relatively accurate (and similar) in how well they predicted someone's party affiliation. On average, they were accurate 67% of the time, and all of the files did a better job at correctly identifying Democrats than Republicans. By contrast, the files varied widely in terms of accurately classifying respondents' education – they ranged from 27% correct to 66% correct depending on the file, and some files were missing education entirely.

You compared what was in the voter files with data from Pew Research Center's American Trends Panel. What finding struck you the most?

The relatively high quality of many of the modeled variables produced by the vendors was an important finding. For example, prior to the 2016 general election, each vendor provided a measure of turnout likelihood in the election. Applying these measures improved the accuracy of the American Trends Panel's estimate of voter preferences in the presidential race. The estimate narrowed Hillary Clinton's advantage from 7 percentage points among all registered voters to a range of 3 to 5 points among likely voters, using the modeled turnout scores. On Election Day, Clinton ended up with a 2-point advantage over Donald Trump in the popular vote. Past voter history is a key component of these models, but the exact algorithms the vendors use aren't public.

Similarly, the vendors' models for predicting someone's partisanship and race were pretty accurate. On average across the five files, *modeled* party affiliation matched *self-reported* party affiliation for about two-thirds of panelists (67%). And 79% of the time, on average, the models available in the voter files accurately predicted the race of our respondents.

You found that the five voter files collectively were more complete and accurate in matching to our American Trends Panel data than any individual file. What lessons should researchers and other users of voter files draw from that?

That's right. Together, the five files we studied were able to match more than nine-in-ten panelists (91%), but the match rates for the individual files were lower, ranging from 50% to 79%. However, each file was able to find people that other files missed, probably because of the different matching algorithms the vendors used.

What we learned in the matching process is that there's a trade-off between matching accuracy (i.e., do you believe the voter file match received is the correct person?) and coverage (i.e., the percent of people able to be matched). Lower match rates generally appeared more accurate, but disproportionately excluded younger and mobile people. High match rates produced more representative samples but may have included more inaccurate matches.

Although we were able to use five different voter files in this research, most researchers will only use a single file. In that case, it's important to think about the goals of your research – depending on what you're investigating, you might value coverage over accuracy or vice versa.

You also found that the files are more reflective of the registered voter population than of the general public as a whole – which makes sense, given the way they are derived. How does that limit their usefulness for researchers and other users?

When voter files first came to prominence for political practitioners and researchers, many were just what the name suggests – lists of registered voters. But as use of voter files for research and targeting has become more widespread, most vendors have tried to cover *all* U.S. adults, including those who aren't registered to vote. Because the core component of the files is a combination of official state lists of registered voters, vendors have sought out commercial databases – available from sources such as credit rating agencies – to locate Americans missing from state voter rolls.

The collective results of the five files provide evidence that the unregistered are not completely invisible to commercial files of the sort examined in this study. Two-thirds of those who told us they were certain that they were not registered were located by at least one of the files. However, researchers should proceed with caution when using this tool to study the general population as a whole, since the kinds of people missed by the files may be very different politically and demographically from the people vendors are able to find.

Given the results of your study, what's your takeaway about how voter files should and shouldn't be used?

Our work doesn't address the use of the files for campaign targeting, which is arguably their most common application. But for researchers, the files offer a relatively easy way to examine the real electorate – who really voted – and to add registration and turnout data to surveys. Campaign pollsters have long relied on the files to provide evidence about past voting as a predictor of future turnout, as well as to provide a sampling frame for election surveys, and we see continued value in that. We haven't yet evaluated the files as sources of samples for surveys of the general public, but stay tuned – we may have more to say on that in the future.

For more on voter files, read the full study, "[Commercial Voter Files and the Study of U.S. Politics.](#)"

Topics Methodological Research, Political & Civic Engagement, Voter Files, Voter Participation

SHARE THIS LINK: SHARE ⬆

Pew Research Center ✳

RESEARCH TOPICS

Politics & Policy

International Affairs

Family & Relationships

Economy & Work

2024/10/30 21:58

Voter files: What are they, how are they used and are they accurate?

Immigration & Migration

Race & Ethnicity

Religion

Age & Generations

Gender & LGBTQ

Science

Internet & Technology

News Habits & Media

Methodological Research

Facebook

X

Tumblr

YouTube

RSS Feed

About   Terms & Conditions   Privacy Policy   Cookie Settings   Feedback   Careers   Contact Us